DOCUMENT RESUME

ED 365 146 FL 021 759

AUTHOR	Oller, John W., Jr.
TITLE	Current Research/Development in Language Testing.
PUB DATE	91
NOTE	32p.; In: Sarinee, Anivan, Ed. Current Developments in Language Testing. Anthology Series 25. Paper presented at the Regional Language Centre Seminar on Language Testing and Language Programme Evaluation (April 9-12, 1990); see FL 021 757.
PUB TYPE	Information Analyses (070) -- Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Analysis of Variance; Language Aptitude; *Language Processing; Language Proficiency; *Language Tests; Language Usage; *Linguistic Theory; Pragmatics; *Predictor Variables; Prior Learning; Research Needs; *Second Languages; Semiotics; Testing; *Test Theory
IDENTIFIERS	*Pierce (Charles Sanders)

ABSTRACT

A discussion of language testing looks at the relationship between the processes of language learning and language testing, particularly from the point of view of pragmatics theory. It outlines some of the theory of Charles Sanders Pierce and its role in the evolution of linguistic theory, as well as the work of other theorists concerning the nature of knowledge and the role of experience in learning. The paper distinguishes between three perspectives in the testing process (those of the text, author, and audience) and examines the tension between them during language testing. It is argued that the three perspectives must be in appropriate correspondence to each other. Research focusing on the three perspectives as they relate to cloze testing is considered; different forms of tests are viewed as focusing on different perspectives. It is concluded that for language testing research and development to be optimally interpretable, the researchers must take care to control the variables of whichever two perspectives are not the focus in the test in question. A brief bibliography is included. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 365 146

CURRENT RESEARCH/DEVELOPMENT IN LANGUAGE TESTING

John W. Oller, Jr

INTRODUCTION

Without question, the most important item on the present agenda for language testing research and development is a more adequate theoretical perspective on what language proficiency is and what sources of variance contribute to its definition in any given test situation. Perhaps the least developed idea with reference to the research has been the differentiation of sources of variance that are bound to contribute to observed differences in measures of language proficiency in different test situations.

Among the sources of variance that have heretofore been inadequately sorted out are those attributable to text/discourse as opposed to authors contrasted also with audience or consumers. With respect to these three positions, which may be roughly related to Peirce's categories of thirdness, firstness, and secondness respectively, several distinct dimensions of each source may be sorted out. Among the most salient variables to be taken into consideration are background knowledge, relative language ability, and motivation of author (first person) and consumer (second person) as well as the properties that can be distinguished as pertaining to the discourse/text itself. For an example or two, these several sources of variability (and others) are discussed within a Peircean perspective relative to research on cloze procedure and several other ways of investigating coherence/comprehensibility of texts/discourses vis a vis certain producers and interpreters. It is argued that impoverished theories that fail to take the three positions of firstness, secondness, and thirdness into consideration are doomed to inadequacy. Nor is research that fails to do so apt to be reasonably interpretable. Examples of experimental research projects that do and do not consider the relevant variables are discussed. Finally, some general recommendations are offered for test development and future research.

FL021759

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Weng

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

GREETING

After ten years, it is a distinct pleasure to be back in Singapore again and to attend once more an international conference at RELC on language testing. As Charles Alderson reminded us at least "a little" has happened in the interim (since the 1980 conference) and we look forward to seeing what the next decade may bring forth. We may hope that all of us who were able to attend this year will be able to come back in ten years time. We are saddened to note that Dr. Michael Canale is no longer with us, and are reminded of our own mortality.

It is a "noble undertaking", as General Ratanakoses (Minister of Education in Thailand and President of SEAMEO) told us yesterday that we are embarked upon, but a difficult one. Therefore, if we are to stay in it for the long haul, as Alderson said, we will require a certain level of "stamina". The Director of RELC, Mr. Earnest Lau and Dr. Jakub Isman, the Director of the SEAMEO Secretariat, defined very admirably at the opening of this year's seminar the scope and limits of the problems that we grapple with and their importance to the enterprise of education especially in multilingual settings. Again and again, in papers at the conference, we are reminded of the central role of language in the communication of information, the establishment and maintenance of social norms, and in the very definition of what education is all about.

A GOAL AND A PLAN

This morning I want to speak to you about current research and development in language testing. Following the recommendation to be "audience-centered", from A. Latief in one of yesterday's sessions, and also a suggestion from Adrian Palmer, I have tried wherever possible to illustrate the various theoretical and practical concerns of my own presentation from things said at the conference. My goal is to introduce a theory of semiosis (our use of the ability we have as human beings to form sensible representations) which regards language testing as a special case. Along the way I will introduce Charles Sanders Peirce [1839-1914], the American scientist, mathematician, logician, and philosopher, best known in this century, perhaps, for having been the mentor of William James and John Dewey.

A GOLDEN RULE FOR TESTERS

In fact, having mentioned Peirce, I am reminded of something he wrote about being audience-centered. By the end of the talk, I hope you will see its relevance to all that I have to say and to the method I have tried to employ in saying it. When he was a young man concerning the process of writing, he wrote in his private journal, "The best maxim in writing, perhaps, is really to love your reader for his own sake" (in Fisch, et al., 1982, p. 9). It is not unlike the rule laid down in the Mosaic law and re-iterated by Christ Jesus that we should love our neighbors as ourselves. It is a difficult rule, but one that every teacher in some measure must aspire to attain. Moreover, in interpreting it with reference to what I will say here today, it is convenient that it may be put in all of the grammatical persons which we might have need of in reference to a general theory of semiosis and to a more specific theory of language testing as a special case.

For instance, with respect to the first person, whether speaker or writer, it would be best for that person to try to see things from the viewpoint of the second person, the listener or reader. With reference to the second person, it would be good to see things (or to try to) from the vantage point of the first. From the view of a third person, it would be best to take both the intentions of the first and the expectations of the second into consideration. And, as Ron MacKay showed so eloquently in his paper at this meeting, even evaluators (acting in the first person in most cases) are obliged to consider the position of "stakeholders" (in the second person position). The stakeholders are the persons who are in the position to benefit or suffer most from program evaluation. They are the persons on the scene, students, teachers, and administrators, so it follows from the generalized version of Peirce's maxim for writers (a sort of golden rule for testers) that evaluators must act as if they were the stakeholders.

Therefore, with all of the foregoing in mind, I will attempt to express what I have to say, not so much in terms of my own experience, but in terms of what we have shared as a community at this conference. May it be a sharing which will go on for many years in a broadening circle of friendships and common concerns. I suppose that our common goal in the "noble undertaking" upon which we have embarked from our different points of view converging here at RELC, is to share our successes and our quandaries in such a way that all of us may benefit and contribute to the betterment of our common cause as communicators, teachers, educators, experimentalists, theoreticians, practitioners, language testers, administrators, evaluators, and what have you.

A BROADER THEORETICAL PERSPECTIVE

It seems that our natural proclivity is to be a little bit cautious about embracing new theoretical perspectives. Therefore, it is with a certain reasonable trepidation that I approach the topic of semiotic theory. Adrian Palmer pointed out that people have hardly had time to get used to the term "pragmatics" (cf. Oller, 1970) before there comes now a new, more difficult and more abstract set of terms drawn from the semiotic theory of Charles Sanders Peirce. It is true that the term "pragmatics" has been at least partially assimilated. It has come of age over the last two decades, and theoreticians around the world now use it commonly. Some of them even gladly incorporate its ideas into grammatical theory. I am very pleased to see that at RELC in 1990 there is a course listed on "Pragmatics and Language Teaching".

Well, it was Peirce who invented the term, and as we press on with the difficult task of sinking a few pilings into solid logic in order to lay as strong a foundation as possible for our theory, it may be worthwhile to pause a moment to realize just who he was.

C. S. Peirce [1839-1914]

In addition to being the thinker who invented the basis for American pragmatism, Peirce did a great deal else. His own published writings during his 75 years, amounted to 12,000 pages of material (the equivalent of 24 books of 500 pages each). Most of this work was in the hard sciences (chemistry, physics, astronomy, geology), and in logic and mathematics. During his lifetime, however, he was hardly known as a philosopher until after 1906, and his work in grammar and semiotics would not become widely known until after his death. His followers, William James [1842-1910] and John Dewey [1859-1952], were better known during their lifetimes than Peirce himself. However, for those who have studied the three of them, there can be little doubt that his work surpassed theirs (see, for example, comments by Nagel, 1959).

Until the 1980s, Peirce was known almost exclusively through eight volumes (about 4,000 pages) published by Harvard University Press between 1931 and 1958 under the title Collected Writings of Charles S. Peirce (the first six volumes were edited by Charles Hartshorne and Paul Weiss, and volumes seven and eight by Arthur W. Burks). Only Peirce scholars with access to the Harvard archives could have known that those eight volumes represented less than a tenth of his total output.

More recently, in 1979, four volumes on mathematics appeared under the editorship of Carolyn Eisele. Peirce's work on mathematics, it is claimed, rivals and surpasses the famed Principia Mathematica by Bertrand Russell and Alfred North Whitehead. In 1982 and 1984 respectively two additional tomes of Peirce's writings have been published by Indiana University Press. The series is titled Writings of Charles S. Peirce: A Chronological Edition and is expected, when complete, to contain about twenty volumes. The first volume has been edited by Max Fisch, et al., (1982) and the second by Edward C Moore, et al., (1984). In his Preface, to the first volume (p. xi), Moore estimates that it would require an additional 80 volumes (of 500 pages each) to complete the publication of the remaining unpublished manuscripts of Peirce. This would amount to a total output of 104 volumes of 500 pages each.

Nowadays even dilettantes (such as Walker Percy a popular writer of novels) consider Peirce to have been a philosopher. In fact, he was much more. He earned his living from the hard sciences as a geologist, chemist, and engineer. His father, Benjamin Peirce, Professor of Mathematics at Harvard was widely regarded as the premier mathematician of his day, yet the work of the son by all measures seems to have surpassed that of the father (cf. Eisele, 1979). Among the better known accomplishments of Charles Sanders Peirce was a mathematical improvement in the periodic table of chemistry. He was also one of the first astronomers to correctly determine the spiral shape of the Milky Way Galaxy. He generalized Boolean algebra - a development which has played an important role in the logic of modern computing. His work in the topological problem of map-making is, some say, still unexcelled.

Ernest Nagel wrote in 1959, "There is a fair consensus among historians of ideas that Charles Sanders Peirce remains the most original, versatile, and comprehensive mind this country has yet produced" (p. 185, also cited by Moore, 1984, p. xi). Noam Chomsky, the foremost linguist and language philosopher of the twentieth century, in an interview with Mitsou Ronat in 1979, said, "The philosopher to whom I feel closest - is Charles Sanders Peirce" (p. 71). In fact, it is Peirce's theory of abduction (or hypothetical inference; see Oller, 1990) that Chomsky credits as the basis for his whole approach to the study of language.

THE CRUCIAL ROLE OF INFERENCE

Peirce himself saw abstract representation and inference as the same thing. Inference, of course, is the process of supposing something on the warrant of

something else, for example, that there will be rain in Singapore because of the build-up of thunderheads all about. Peirce wrote, "Inference in general obviously supposes symbolization; and all symbolization is inference. For every symbol ... contains information. And ... all kinds of information involve inference. Inference, then, is symbolization. They are the same notions" (1865, in Fisch, 1982, p. 280). The central issue of classic pragmatism, the variety advocated by Peirce, was to investigate "the grounds of inference" (1865, in Fisch, p. 286), or, in different words, the connection of symbols and combinations of them with the world of experience. However, Peirce differed from some so-called "pragmatists" because he did not see experience as supplying any basis for inference, but rather, inference as the only possible basis for experience. In this he was encouraged by his precursor Immanuel Kant, and his position would be later buttressed by none other than Albert Einstein (see pertinent writings of Einstein in Oller, 1989).

PRAGMATIC MAPPING

Figure 1 gives a view of what I term "pragmatic mapping". It is by definition the articulate linking of text (or discourse) in a target language (or in fact any semiotic system whatever), with facts of experience known in some other manner (i.e., through a different semiotic system or systems).

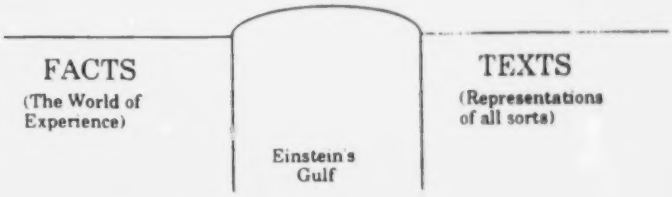

Figure 1. Pragmatic mapping.

That is, pragmatic mapping (also known as abductive reasoning), is a kind of translation process. It is a process of taking a representation in one form and interpreting it in terms of a representation in some other form. The only thing that keeps this process from being completely circular, and therefore empty, is that we really do have some valid knowledge of facts in an external world. Another point to be made is that the process of pragmatic mapping also involves risk. Or as James Pandian put it at this conference, "We talk a lot about what we don't know". Or putting the point in a slightly weaker form, we only have some of the facts most of the time and we are seeking to discover others or we may merely be speculating about them.

THE PLACE FOR SKEPTICISM

To some extent, therefore, British skepticism of the sort advocated by David Hume [1711-1776] and Bertrand Russell [1872-1970] was only partially well-founded. If there were no secure knowledge, and if all representations were always of doubtful interpretation in all circumstances (which they are not), then all representations would ultimately be meaningless, and communication and language acquisition would be impossible. However, both communication and language acquisition do in fact occur, and are in fact possible precisely because we do possess a great deal of well-equilibrated knowledge (previously established pragmatic mappings) concerning the external world--a world that is as real as the space-time continuum can be. All of this is thrashed out in detail in Oller (1989) through a collection of writings by Einstein, Peirce, James, de Saussure, Russell, Dewey, and Piaget, so that argument will not be reiterated here. Let it simply be noted that for all of its merits in pointing out the naiveness of naive realism and the positive benefits of empiricism, British skepticism failed to so much as touch the skin of classic pragmatism or the Peircean idea of abductive reasoning which forms the basis for the diagram given in Figure 1.

There are two interpretations of the figure that are of interest here. First, there is the general theory that it suggests for the comprehension of semiotic material, i.e., texts or discourse, in general, and second, there is the more specific application of it to language testing theory which we are about to develop and elaborate upon.

NECESSARY AND SUFFICIENT CONDITIONS

With respect to the first interpretation we may remark that the theory of pragmatic mapping, though entirely neglected by reviewers like Skehan (1989), offers both the necessary and sufficient conditions for language comprehension and acquisition. In order for any individual to understand any text it is necessary for that individual to articulately map it into his or her own personal experience. That is, assuming we have in mind a particular linguistic text in a certain target language, the comprehender/acquirer must determine the referents of referring noun phrases (who, what, where, and the like), the deictic significances of verb phrases (when, for how long, etc.), and in general the meanings of the text. The case is similar with the producer(s) of any given text or bit of text. All of the same connections must be established by generating surface forms in a manner that articulately corresponds to facts. If such texts are comprehended and produced (here I diverge from Krashen somewhat) over a sufficient period of time, the outcome is language acquisition. For this to occur, it figures that the individual in question must both have access to comprehensible input and must engage in comprehending it. Moreover, the learner must actively (productively) engage in the articulate linking of texts in the target language with his or her own experience. In fact, comprehension already entails this much even before any active speaking or writing ever may take place. This entails sufficient motivation in addition to opportunity. Therefore, the theory of pragmatic mapping provides both the necessary and sufficient conditions for language acquisition (whether primary or non-primary).

EINSTEIN'S GULF

Obviously, the theory requires elaboration. Before going on to a slightly elaborated diagram viewing the process in terms of a hierarchy of semiotic capacities, however, a few comments are in order concerning the middle term of Figure 1 which is referred to as "Einstein's gulf". Although it may be true that there really is an external world, and though we may know quite a lot about it (albeit practically nothing in relation to what is to be known; see the reference to Pandian above), our knowledge of the world is always in the category of being an inference. There is no knowledge of it whatever that does not involve the inferential linking of some representational form (a semiotic text of some sort) with the facts of experience. The physical world, therefore, the cosmos in all its

vast extent, we do not know directly--only indirectly and inferentially through our representations of it.

The fact that physical matter should be representable at all is as Einstein put it, miraculous. He wrote of a "logically unbridgeable gulf" which "separates the world of sensory experiences from the world of concepts and propositions" (Einstein, 1944, in Oller, 1989, p. 25). This gulf poses an insurmountable barrier to any theory that would attempt to explain human intellect in a purely materialistic manner. All materialistic philosophies end in the abyss. There is for them, no logical hope whatever. It would be good to dwell on the philosophical and other implications of this, but we cannot linger here.

FACTS ARE INDEPENDENT OF SOCIAL CONSENSUS

Another point worthy of a book or two, is that what the material world is, or what any other fact in it is, i.e., what is real, in no way depends on what we may think it to be. Nor does it depend on any social consensus. Thus, in spite of the fact that our determination of what is in the material world (or what is factual concerning it), is entirely dependent on thinking and social consensus (and though both of these may be real enough for as long as they may endure), reality in general is entirely independent of any thinking or consensus. Logic requires, as shown independently by Einstein and Peirce (more elaborately by Peirce), that what is real must be independent of any human representation of it. But, we cannot develop this point further at the moment. We must press on to a more elaborate view of the pragmatic mapping process and its bearing on the concerns of language testers and program evaluators.

APPLIED TO LANGUAGE TESTING

In fact, the simplest form of the diagram, Figure 1, shows why language tests should be made so as to conform to the naturalness constraints proposed earlier (Oller, 1979, and Doye, this conference). It may go some way to explaining what Read (1982, p. 102) saw as perplexing. Every valid language test that is more than a mere working over of surface forms of a target language must require the linking of text (or discourse) with the facts of the test taker's experience. This was called the meaning constraint. The pragmatic linking, moreover, ought to take place at a reasonable speed--the time constraint. In his

talk at this conference, Alderson stressed, as others have throughout, the importance of reliability and validity. It is validity that the naturalness constraints are concerned with directly.

THE SEMIOTIC HIERARCHY

Figure 2 gives a more developed view of the pragmatic mapping process. As my point of reference here at this year's RELC seminar for what follows immediately, I take N. F. Mustapha's suggestion, that we must look at the psycho-motor functions that enter into the taking of a language test.

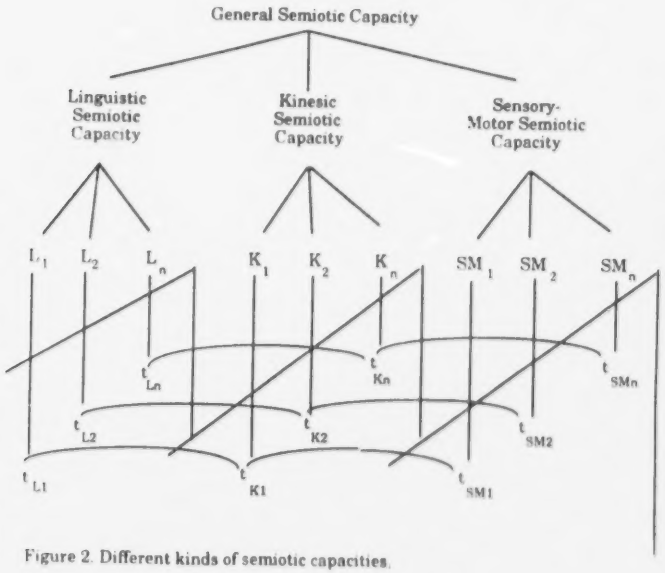

Figure 2. Different kinds of semiotic capacities.

The new diagram, therefore, suggests that a hierarchical organization exists. At the top of the hierarchy is what might be called general semiotic capacity. This is our ability to represent facts at the highest level of abstraction imaginable. It undergirds all the less general and more specialized capacities by which we make sense of our world. At the next level down we find at least three (perhaps there are more, but there cannot be any less) universal human capacities that are also of a representational (semiotic) sort: linguistic, kinesic, and sensory-motor. In their most abstract and general forms, each of these capacities is nonetheless distinct. Linguistic ability is the one most studied by us language testers so we may pass over it for the moment.

Kinesic Capacity. Kinesic ability pertains to our knowledge of the meanings of gestures, some aspects of which are universal and some of which are conventional and must be acquired. Smiling usually signifies friendliness, tears sadness, and so on, though gestures such as these are always ambiguous in a way that linguistic forms are not ordinarily. A smile may be the ultimate insult and tears may as well represent joy as sorrow. Sensory-motor representations are what we obtain by seeing, hearing, touching, tasting, and smelling. They include all of the visceral and other sensations of the body.

Sensory-Motor Capacity. Sensory-motor representations, as we learn from empiricism, are the starting point of all experience, experimentation, and therefore of science, and yet a little logic soon reveals that they are insufficient to determine anything by themselves (this was the valid point to be derived from the skepticism of Hume and Russell, see Oller, 1989 for elaboration). The problem with sensory-motor representations is to determine what precisely they are representations of. What do we see, hear, etc? The general logical form of the problem is a Wh-question with an indeterminate but emphatic demonstrative in it: namely, "What is *that*?" To see the indeterminacy in question, picture a scientist in a laboratory with a surprised expression on his face looking at a strange new concoction in a test-tube, or under a microscope, on a CRT, or in a mathematical formula, or wherever, and asking, "What is that?" Or imagine a person on the street or a language tester who asks the same question of any observed datum.

A gesture may help the observer determine whatever is in question. For instance, if someone points to whatever is in question or merely looks at it, this narrows down the field of possible resolutions of the demonstrative reference, but it never can adequately determine the phenomenon or object in question unless it is supported by something more abstract--namely, a conceptual or linguistic representation. With the gesture alone there is always the problem of finding out what it refers to. What precisely is pointed to or signified? In

BEST COPY AVAILABLE

experience, gestures may serve deictic or other significant functions, but, as Peirce pointed out, gestures are always reactionally degenerate. Sensory-motor representations are also degenerate, but in a rather different way. They actually fade or dissipate over time, or even if they can be well-preserved, the physical facts themselves to which the sensory-motor impressions correspond will change and thus distort the connection between the sensory-motor representation and whatever it purports to represent.

Linguistic Capacity. Here is where language comes to the rescue. While sensory-motor representations by themselves are entirely inadequate to determine any facts about experience completely, and gestures hardly help except to bring certain significances to our attention, language affords the kind of abstract conceptual apparatus necessary to fully determine many of the facts of experience. For instance, it is only by linguistic supports that we know that today we are in Singapore, that it is Tuesday, April 9, 1990, that Singapore is an island off the southern tip of Malaysia, and west of the Philippines and north of Australia, that my name is John Oller, that Edith Hanania, Margaret Des Brisay, Liz Parkinson, Jagjeet Singh, Ron MacKay, Adrian Palmer, Kanchana Prapphal, P. W. J. Nababan, James Pandian, Tibor von Elek, and so forth, are in the audience. We know who we are, how we got to Singapore, how we plan to leave and where we would like to go back to after the meeting is over, and so forth. Our knowledge of all of these facts is dependent on linguistic representations. If any one of them were separated out from the rest, perhaps some reason could be found to doubt it, but taken as a whole, the reality suggested by our common representations of such facts is not the least bit doubtful. Anyone who pretends to think that it is doubtful is in a state of mind that argumentation and logic will not be able to cure. So we will pass on.

Particular Systems and Their Texts. Beneath the three main universal semiotic capacities identified, various particular systems are indicated. Each of these requires experience and acquisition in order to connect it to the class of texts which it defines. Each specialized semiotic system, it is asserted, supertends, or defines (in the manner of a particular grammatical system), a class of texts, or alternatively, is defined in part by the universal system that underlies it and in part by the texts that it relates to.

Relevance to Language Testing Illustrated. Now, let's see how this hierarchical model is relevant to language testing. John Read, in his very informative paper, without perhaps intending to, showed the relevance of several aspects of this model. For instance, one of the critical aspects of language use in the writing process is not merely language proficiency per se, which is represented as any given L_i in the diagram, but is also dependent on background knowledge which may have next to nothing to do with any particular L_i. The

background knowledge can only be expressed representationally as some combination of linguistic, gestural (especially indexical signs), and sensory-motor representations. It is at least obtained through such media. Perhaps in its most abstract form it is represented in purely abstract logical forms, at least part of whose structure, will be propositional in character (i.e., equilibrated relations between subjects and predicates, negations of these, and concatenations of various conjunctive and disjunctive sorts). However, knowledge which is not ultimately grounded in or related to sensory-motor contexts (i.e., sensory-motor representations) is mere superstition or pure fiction. That sort of knowledge we can know nothing of because it has no bearing on our experience.

THREE SORTS OF RESULTS PREDICTED

Looking at the pragmatic mapping process in terms of the proposed hierarchy predicts three kinds of results of immediate importance to us language testing researchers and program evaluators. Each sort of result is discussed in one way or another in papers at this conference, and it may be useful to consider each in turn.

(i) **Distinct Factor(s) Explained.** As John Read, Achara Wongsatorn, and Adrian Palmer showed, language proficiency can be broken into a variety of factors and, as Read argued most convincingly, language proficiency per se can properly be distinguished (at least in principle) from background knowledge. Each of the various factors (sometimes trait, sometimes skill, and sometimes method) involves different aspects of the hierarchy. For example, this can easily be demonstrated logically (and experimentally as well) with respect to the distinctness of background knowledge from language proficiency by seeing that the same knowledge can be expressed more or less equivalently in L_1, L_2, or in fact in any L_i whatever that may be known to a given user or community of users. Therefore, background knowledge is distinct from language proficiency.

(ii) **General Factor(s) Explained.** However, the hierarchical view of the theory of pragmatic mapping also shows that background knowledge and language proficiency must be inevitably interrelated. This is logically obvious from the fact that the theory (following Peirce) asserts

that all comprehension and all representation is accomplished via a complex of translation processes. That is to say, if each and every semiotic representation must be understood by translating it into some other form, it follows that the various forms must have some common ground. The hypothesizing of "general semiotic capacity" at the deepest level of the hierarchy expresses this fact most perfectly, but, in fact, every node in the hierarchy suggests the interrelatedness of elements above and below that node. Hence, we have a fairly straightforward explanation for the generally high correlations between language proficiency, school achievement, IQ tests, subject matter tests, as well as the interdependency of first and second language proficiency, and many similar interactions. The general factor (more likely, factors, as John Carroll has insisted) observed in all kinds of educational or mental testing can be explained in this way.

(iii) **Non-Linearity Predicted**. The interrelatedness of elements in the hierarchy, furthermore, is bound to increase with increasing maturity and well-roundedness of experience, i.e., at higher and better integrated levels of experience. This result has been commented at this year's RELC seminar by Charles Stansfield in public discussion with Alderson (also see Oltman, Stricker, and Barrows, 1990). We see in a straightforward way why it is that as normal human beings mature, skills in all the various elements of the semiotic hierarchy are bound to mature at first at rather different rates depending on experience. This will produce, in the early stages, rather marked differences in basic skills (Figure 3) and traits (or components of language proficiency, Figure 4), just as Palmer pointed out at this seminar with reference to the sort of model that Canale and Swain, and Palmer and Bachman have argued for.

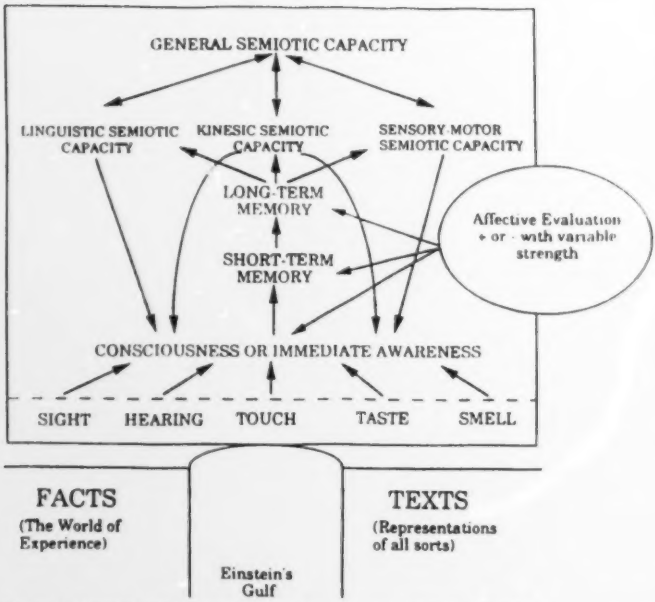

Figure 3. A modular information processing expansion of the pragmatic mapping process.

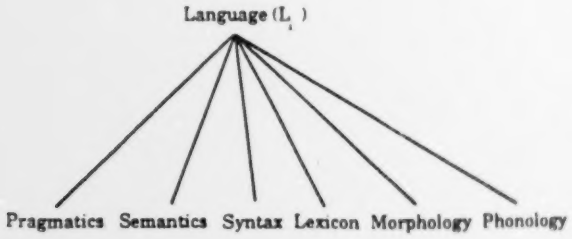

Figure 4. Language proficiency in terms of domains of grammar.

However, as more and more experience is gained, the growth will tend to fill in gaps and deficiencies such that a greater and greater degree of convergence will naturally be observed as individuals conform more and more to the semiotic norms of the mature language users of the target language community (or communities). For example, in support of this general idea, Oltman, Stricker, and Barrows (1990) write concerning the factor structure of the Test of English as a Foreign Language that "the test's dimensionality depends on the examinee's overall level of performance, with more dimensions appearing in the least proficient populations of test takers" (p. 26). In addition, it may be expected that as maturation progresses, for some individuals and groups, besides increasing standardization of the communication norms, there will be a continuing differentiation of specialized subject matter knowledge and specialized skills owing to whatever differences in experience happen to be sustained over time. For example, a person who speaks a certain target language all the time will be expected to advance in that language but not in one that is never experienced. A person who reads lots of old literary works and studies them intently is apt to develop some skills and kinds of knowledge that will not be common to all the members of a community. Or, a person who practices a certain program of sensory-motor skill, e.g., playing racquetball, may be expected to develop certain skills that a marathoner will not necessarily acquire, and so forth throughout the limitless possibilities of the hierarchy.

An Information Processing View. Another way of looking at the same basic hierarchy of semiotic capacities, still in relation to the pragmatic mapping theory, is in terms of information processing, as shown in Figure 5.

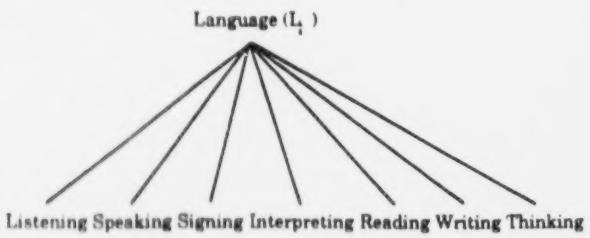

Figure 5. Language proficiency in terms of modalities of processing.

Here the general question is what sorts of internal processing go on as a language user either produces or interprets representations in relation to facts of experience. The more specific question, of interest to language testing, is how does the test taker relate the text (or discourse) of the test to the facts of his or her own experience. The general outlines of the model may be spelled out as follows. Information impinges on the language user from the external world first through the senses. We might say that this is the first line of defense, and it feeds directly into consciousness or immediate awareness. At the same time consciousness is also guided by expectations coming from the various internalized grammatical systems, linguistic, kinesic, and sensory-motor. As information is processed according to these several inter-coordinated, and to some extent co-dependent expectancy systems, what is understood passes to short-term memory while whatever is not understood is filtered out as-it-were, even though it may in fact have been perceived. What is processed so as to achieve a deep level translation into a general semiotic form goes into long term memory. All the while information being processed is also evaluated affectively for its content, i.e., whether it is good (from the vantage point of the processor) or bad. In general, the distinction between a positive or negative marking, and the degree of that markedness, will determine the amount of energy devoted to the processing of the information in question. Things which are critical to the survival and well-being of the organism will tend to be marked positively in terms of affect and their absence will be regarded negatively.

Affect as Added to Cognitive Effects. The degree of importance associated with the object (in a purely abstract and general sense of the term "object") will be determined by the degree of positive or negative affect associated with it. To some extent this degree of markedness and even whether a given object of semiosis is marked positively or negatively will depend on voluntary choices made by the processor. However, there will be universal tendencies favoring survival and well-being of the organism. This means that on the positive side we will tend to find objects that human beings usually regard as survival enhancing and a complementary set of negative elements that will usually be seen as undesirable.

With respect to language processing more specifically, the consequences of affective evaluation are immense. We know of many experimental effects which show both the importance of positive and correct cognitive expectancies (these presumably from the semiotic hierarchy of capacities: linguistic, kinesic, and sensory-motor) and of positive or negative affective valuations of objects of perception, awareness, and memory. These effects are sometimes dramatic and relatively easy to illustrate. In tachistoscopic presentations of stimuli, it is well-

known that contextually expected words, for instance, are easier to perceive than unexpected ones (the British psychologist John Morton comes to mind in this connection). In fact, either positive or negative expectations may be created by context which either make it easier or in fact make it harder than average to perceive a given item. These experiments carry over rather directly into the whole genre of cloze testing to which we will return shortly. However, it can be demonstrated that in addition to the effects of cognitive expectancies, affective evaluations associated with stimuli also have additional significant and important (a distinction made by James Dean Brown [1988] and alluded to by Palmer at this meeting) effects on processing. For instance, when we are hearing a conversation amid background noise and not listening, we are apt to perk up our ears so-to-speak whenever we hear our own name mentioned. It is as if the ears themselves were specially tuned for the mention of our own name. This effect and others like it, well-known to experimental psychologists are collectively known under the terms perceptual vigilance and perceptual defense. The latter phenomenon is common to the difficulty we sometimes experience in perceiving something we really don't want to see (e.g., obscenities or representations pertaining to death, and the like).

Relating all of the foregoing to language testing, I am reminded again of Read's paper of yesterday evening. As he pointed out the evidence seems to suggest that writers who are highly motivated and well-informed do better on all sorts of writing tasks. They generally write more, at a greater level of complexity, and with greater coherence. Furthermore, the graders and anyone else who takes the time to read such essays find the ones written by better motivated and better informed writers to also be that much more comprehensible. All of which leads me to the most important and final diagram for this paper, Figure 6.

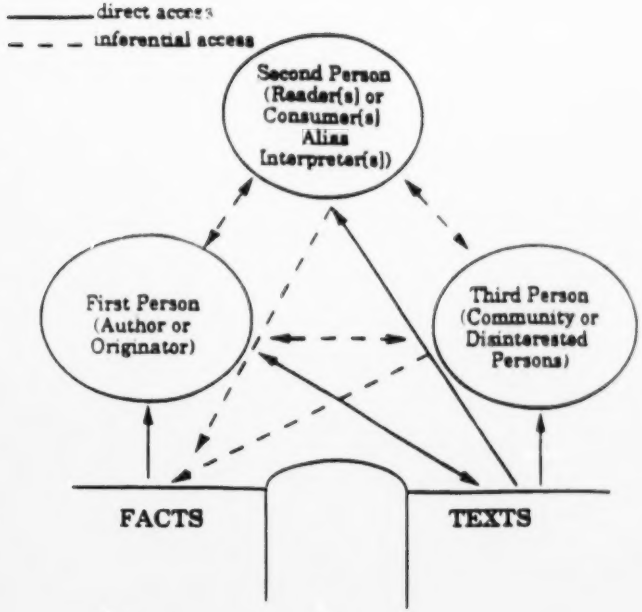

Figure 6. The three Peircean categories as positions or perspectives of persons in reference to cloze test performances (dotted lines indicate indirect inferential connections, while solid lines indicate more or less direct perceptual connections).

FIRST, SECOND, AND THIRD PERSPECTIVES

Not only is it necessary in language testing research and in program evaluation to develop a more comprehensive and better defined theoretical perspective on what semiotic capacities and processes there are, and how they interrelate with each other, but it is also, I believe, urgently necessary to differentiate the various perspectives of the persons involved in the process. The first person or producer of discourse (or text) is obviously distinct from the second person or consumer. What is not always adequately appreciated, as

BEST COPY AVAILABLE

Read points out in his paper at this meeting, is that variability in language tests may easily be an indiscriminant mix from both positions when only one is supposedly being tested. What is more, logically, there is a third position that is shared by the community of users (who will find the text meaningful) and the text itself. Incidentally, for those familiar with Searle's trichotomy in speech act theory (a rather narrow version of pragmatic theory), we may mention that what he calls illocutionary force (or meaning) pertains to the first position, perlocutionary force to the second, and mere locutionary force to the third.

It will be noted that the first person is really the only one who has direct access to whatever facts he or she happens to be representing the production of a particular text. Hence, the first person also has direct access to the text. At the same time the text may be accessible directly to the person to whom it is addressed, but the facts which the text represents (or purports to represent in the case of fiction) are only indirectly accessible to the second person through the representations of the first. That is, the second person must infer the intentions of the first person and the facts (whatever either of these may be). Inferences concerning those facts are based, it is hypothesized, on the sort of semiotic hierarchy previously elaborated (Figures 1-5). Similarly, a third person has direct access neither to the facts nor the intentions of the first person nor the understandings of them by the second person. All of these points must be inferred, though the text is directly accessible. The text, like the third person(s), also logically is part of the world of facts from the point of view of the third person, just as the first person and second person are part of that world. (For anyone who may have studied Peirce's thinking, the three categories differentiated here will be readily recognized as slightly corrupted, i.e., less abstract and less general, versions of his perfectly abstract and general categories of firstness, secondness, and thirdness.)

Going at these categories in a couple of different ways, I am sure that I can make clearer both what is meant by them in general and how they are relevant to the practical business of language testing. When, as language testers, we ask questions about skills and traits, as Canale and Swain (see Palmer's references) did and as Palmer and Bachman have in their several joint projects (again, see Palmer's references), we are concerned primarily in most cases with what is going on in either the first or second position. However, with some procedures attention shifts to the third position, e.g., when we use language tests to investigate characteristics of textual structure.

The point that I want to make in this next section is that unless the two other positions (beyond whichever of the three may already be in focus), and possibly a great many subtle variables within each, are controlled, it is likely that

data drawn from any language testing application will be relatively meaningless. Unfortunately this is the case with far too many studies. As Palmer emphasized in his review of program evaluation with respect to theories of language acquisition and whatever sorts of proficiency may be acquired, it appears that the language teaching profession is long on methods, recipes, and hunches, and short on theories that are clear enough to put to an experimental test.

TESTING PROCEDURES AS PROVING GROUNDS

For instance, consider cloze procedure as a family of testing techniques. Between 1983 and the end of 1989 about 717 research projects of a great variety of sorts were conducted using cloze procedure in one way or another. A data search turned up 192 dissertations, 409 studies in ERIC, and 116 in the PsychLit database. At this conference there were a number of other studies that either employed or prominently referred to cloze procedure (but especially see R. S. Hidayat, S. Boonsatorn, Andrea Penaflorida, Adrian Palmer, David Nunan, and J. D. Brown). We might predict that some of the many cloze studies in recent years, not to mention the many other testing techniques, would focus on the first person position, i.e., variability attributable to the producer(s) of a text (or discourse); some on the second person position, variability attributable to the consumer(s); and some on third position, variability attributable to the text itself. Inevitably, studies of the third position relate to factors identified with a community of language users and the sorts of texts they use.

Always a Tensional Dynamic. In fact, the interaction between a writer (or speaker) and a reader (or listener) through text (or discourse) is always a dynamic tensional arrangement that involves at least three positions simultaneously. Sometimes additional positions must be posited, but these, as Peirce showed, can always be seen as complications of the first three positions. All three of the basic positions also logically entail all of the richness of the entire semiotic hierarchy elaborated previously in this paper (Figures 1-5). Also, as John Read hinted (and as Peter Doyé stated overtly), we may move the whole theory up a level of abstraction and consider that "test-raters are different people from the test-makers, and that the way the raters interpret the task is a further source of variability in the whole process" (Read, this conference). What is not apparent in Read's statement, though I don't think he would deny it, is that the problem hinted at is completely general in language testing research and applications. All tests are susceptible to the same sort of logical criticism in

terms of the sources of variability that will influence scores on them.

Congruence or Goodness-of-Fit as the Central Issue. In effect, the question throughout all the levels of abstraction that are imaginable, as Doyé correctly intuited though he did not say this explicitly, is whether or not the various possible positions of interlocutors (first and second positions) and texts (third), testers (first position once removed) and tests (third position once removed) interlocutors and texts, raters (first position twice removed) and testers and interlocutors and texts, etc., are in agreement. It is miraculous (as Einstein observed decades ago, see Oller, 1989) that any correspondence (i.e., representational validity) should ever be achieved between any representations and any facts, but it cannot be denied that such well-equilibrated pragmatic mappings are actually common in human experience. They are also more common than many skeptics want to admit in language testing research as well, though admittedly the testing problem is relatively (and only relatively) more complex than the basic communication problem. However, I believe that it is important to see that logically the two kinds of problems are ultimately of the same class. Therefore, as testers (just as much as mere communicators) we seek convergences or "congruences" (to use the term employed by Peter Doye) between tests and what they are supposed to be tests of.

Reality and even authenticity (apart from the idea of congruence as defined within the theory of pragmatic mapping or the correspondence theory of truth which is the same thing; cf. Oller, 1990), on the other hand, are hardly worth discussing since they are so easy to achieve in their minimal forms as to be trivial and empty criteria. Contrary to a lot of flap, classrooms are real places and what takes place in them is as real as what takes place anywhere else (e.g., a train station, restaurant, ballpark, or you name it!) and to that extent tests are as real and authentic in their own right as any other superficial semiotic event. Interviews are real enough. Conversations, texts, stories, and discourse in general can be just as nonsensical and ridiculous outside the classroom (or the interview, or whatever test) as in it. Granted we should get the silliness and nonsense out of our teaching and our testing and out of the classroom (except perhaps when we are merely being playful which no doubt has its place), but reality and authenticity apart from a correspondence theory of truth, or the pragmatic mapping theory outlined here, are meaningless and empty concepts.

Anything whatever that has any existence at all is ipso facto a real and authentic fact. Therefore, any test no matter how valid or invalid, reliable or unreliable, is ipso facto real and, in this trivial way, authentic. The question is whether it really and authentically corresponds to facts beyond itself. But here we introduce the whole theory of pragmatic mapping. We introduce all of Peirce's theory of abduction, or the elaborated correspondence theory of truth.

The test is seen as representative of something else. It is the correspondence to that something else which is really at issue. We introduce the matter of validity, truth, and goodness of fit in relation to an external world beyond the test per se. Tests, curricula, classrooms, teachers and teaching are all real enough, the problem is to authenticate or validate them with reference to what they purport to represent.

With reference to that correspondence issue, without going into any more detail than is necessary to the basic principles at stake let me refer to a few studies that show the profound differences across the several pragmatic perspectives described in Figure 6. Then I will reach my conclusion concerning all of the foregoing and hopefully justify in the minds of participants in the conference and other readers of the paper the work that has gone into building up the entire semiotic theory in the first place. There are many examples of studies focussing on the first position, though it is the least commonly studied position with cloze procedure. A dramatically clear example is a family of studies employing cloze procedure to discriminate speech samples drawn from normals from samples drawn from psychotics.

The First Person in Focus. When the first person is in focus, variability is attributable to the author (or speaker) of the text (or discourse) on which the cloze test is based. In one such study, Maher, Manschreck, Weinstein, Schneyer, and Okunieff (1988; and see their references), the third position was partially controlled by setting a task where the subjects described Breughel's "The Wedding Feast". Then cloze tests were made by replacing every fifth word with a standard blank. Paid volunteers (n = 10), then, were asked to "rate" (i.e., fill in the blanks on the various) speech samples with a minimum of two raters per sample. The assumption here being that the second position variability will be negligible. (In fact, this assumption will turn out to be wrong in this case just as it so often is in others). Results then were pooled across raters and the various authorial groups were contrasted. In fact, some discrimination did appear between different samples of speech, but (and this is the critical point to our theory), the researchers realized rather late that the second position involved variables that might drastically affect the outcomes.

A follow up study in fact aimed to test whether more educated "raters" (i.e., the paid volunteers who filled in the cloze tests) might be better at guessing all kinds of missing items and therefore might produce a ceiling effect. In such a case any differences between the speech samples of normals and psychotics would be run together at the top of the scale and thereby washed out. Indeed the follow up confirmed this expectation and it was concluded that less educated (and probably, therefore, less proficient) "raters" would generally produce greater discrimination among normal and psychotic speech samples. In addition

to demonstrating that cloze procedure is sensitive to differences in the first position for psychotics and normals, this study (albeit unintentionally) showed how the procedure has to be tuned to the right level of difficulty for "raters" (i.e., persons in the second position) in order to get results. Another alternative would have been to adjust the level of difficulty of the task performed by the normals and psychotics thereby producing more complex passages (in the third position) to be cloze-rated.

Another pair of studies that focussed on first position variability with cloze procedure sought to differentiate plagiarists from students who did their own work in introductory psychology classes. In their first experiment (E1), Standing and Gorassini (1986) showed that students received higher scores on cloze passages over their own work (on an assigned topic) than over someone else's. Subjects were 16 undergraduates in psychology. In a follow-up with 22 cases, E2, they repeated the design but used a "plagiarized" essay on a new topic. In both cases, scores were higher for psychology students who were filling in blanks on their own work.

Clearly the researchers assumed in both E1 and E2 that they had sufficiently controlled the variability attributable to differences in the second position, i.e., that of the subject filling in the blanks on one or another cloze passage, and in the third, i.e., the text itself. The researchers assumed that the texts in E1 would be reasonably comparable since they were all written on an assigned topic. John Read's paper at this meeting shows that in many cases this assumption will probably not be correct. In fact, it seems fairly likely that a really bright plagiarist, one who knew the subject-matter well and who was highly proficient in the language at issue in the plagiarized material, might very well escape detection. Motivation of the writers, the amount of experience they may have had with the material, and other background knowledge are all uncontrolled variables.

With respect to E2, the third position is especially problematic. Depending on the level of difficulty of the text selected, it is even conceivable that it might be easier to fill in the blanks in the "plagiarist's" work (the essay from an extraneous source) than for some subjects to recall the exact word they themselves used in a particularly challenging essay. There is also a potential confounding of first and second positions in E1 and in E2. Suppose one of the subjects was particularly up at the time of writing the essay and especially depressed, tired, or down at the time of the cloze test. Is it not possible that an honest student might appear to be a plagiarist? Or vice versa? At any rate, difficulty, topic, level of abstraction, vocabulary employed, motivation, alertness, and a host of other factors that might be present at the time of writing and not at the filling in of the blanks (or vice versa) are potential confounding variables.

Nevertheless, there is reason to hold out hope that under the right conditions cloze procedure might be employed to discourage if not to identify plagiarists, and it should be obvious that countless variations on this theme, with reference to the first position, are possible.

The Second Person in Focus. As an example of a study focussing on the second position, consider Zinkhan, Locander, and Leigh (1986). They attempted to determine the relative effectiveness of advertising copy as judged by recallability. Two independent dimensions were identified: one affective, relating to how well the subjects (n = 420) liked the ad, brand, and product category, and one cognitive relating to knowledge and ability of the subjects (we may note that background knowledge and language proficiency are confounded here but not necessarily in a damaging way). Here, since the variability in advertising copy (i.e., third position) is taken to be a causal factor in getting people to remember the ad, it is allowed to vary freely. In this case, the first position effectively merges with the third, i.e., the texts to be reacted to. It is inferred then, on the basis of the performance of large numbers of measures aimed at the second position (the n of 420), what sorts of performances in writing or constructing ads are apt to be most effective in producing recall. In this instance since the number of cases in the second position is large and randomly selected, the variability in second position scores is probably legitimately employed in the inferences drawn by the researchers as reflecting the true qualitative reactions of subjects to the ads.

Many, if not most, second language applications of cloze procedure focus on some aspect of the proficiency or knowledge of the reader or test taker. Another example is the paper by R. S. Hidayat at this conference who wrote, "Reading as a communicative activity implies interaction between the reader and the text (or the writer through the text). To be able to do so a reader should contribute his knowledge to build a 'world' from information given by the text." I would modify this statement only with respect to the "world" that is supposedly "built" up by the reader (and or the writer). To a considerable extent both the writer and the reader are obligated to build up a representation (on the writer's side) and an interpretation (a representation of the writer's representation, on the reader's side) that conforms to what is already known of the actual world that reader, writer, and text are all part of (in defense of this see the papers by Peirce, Einstein, Dewey, and Piaget in Oller, 1989). In an even more important way, the reader's interpretation should conform in some degree to the writer's intended meaning, or else we could not say that any communication at all had occurred. Therefore, the reader had better aim to build just the world that the writer has in mind, not merely some "possible world" as so many theoreticians are fond of saying these days. Similarly, the writer, unless he or she is merely

building up a fictional concoction had best have in mind the common world of ordinary experience. Even in the case of fiction writing, of course, this is also necessary to a very great extent, or else the fiction will become incomprehensible.

Happy to say, in the end, Hidayat's results are completely in line with the theory advocated here. They show a substantial correlation between the several tests aimed at grammar, vocabulary, and whatever general aspects of comprehension are measured by cloze. This is as we should expect, at least for reasonably advanced learner/acquirers. Witness prediction (ii) above that as language learners mature towards some standard level their various skills and components of knowledge will tend more and more to even out and thus to be highly correlated--producing general semiotic factors in correlational research. This being the case, apparently, we may conclude that the first and third positions were adequately controlled in Hidayat's study to produce the expected outcome in the second position.

In addition, relative to observed general factors in language testing research, recall (or refer to) the high correlations reported by Stansfield at this conference. His results are doubly confirmatory of the expected convergence of factors in the second position for relatively advanced learners (see prediction ii above) because, for one, he used a pair of rather distinct oral testing procedures, and for two, he did it with five replications using distinct language groups. In Stansfield's case, the oral tests, an Oral Proficiency Interview (OPI) and a Simulated Oral Proficiency Interview (SOPI), are themselves aimed at measuring variability in the performance of language users as respondents to the interview situation--i.e., as takers of the test regarded as if in second position. Though subjects are supposed to act as if they were in first position, since the interview is really under the control of the test writer (SOPI) or interviewer (OPI), subjects are really reactants and therefore are seen from the tester's point of view as being in second position. As Stansfield observes, with an ordinary OPI standardization of the procedure depends partly on training and largely on the wits of the interviewer in responding to the output of each interviewee.

That is to say, there is plenty of potential variability attributable to the first position. With the SOPI, variability from the first position is controlled fairly rigidly since the questions and time limits are set and the procedure is more or less completely standardized (as Stansfield pointed out). To the extent that the procedure can be quite perfectly standardized, rater focus can be directed to the variability in proficiency exhibited by interviewees (second position) via the discourse (third position) that is produced in the interview. In other words, if the first position is controlled, variability in the third position can only be the responsibility of the person in second position.

With the OPI, unlike the case of the SOPI, the interviewer (first position) variability is confounded into the discourse produced (third position). Therefore, it is all the more remarkable when the SOPI and OPI are shown to correlate at such high levels (above .90 in most cases). What this suggests is that skilled interviewers can to some extent factor their own proficiency out of the picture in an OPI situation. Nevertheless, cautions from Ross and Berwick (at this conference) and Bachman (1988) are not to be lightly set aside. In many interview situations, undesirable variability stemming from the first position (the interviewer or test designer) may contaminate the variability of interest in the second position. This caveat applies in spades to variability with respect to particular individuals interviewed though less so as the number of interviewees is increased. To avoid undesirable contamination from the first position, the interviewer (or test writer) must correctly judge the interests and abilities of the interviewee in each case so as not to place unnecessary stumbling blocks in the way. Apparently this was accomplished fairly successfully on the whole (though one wonders about individual cases) in Stansfield's study or else there would be no way to account for the surprisingly strong correlations between OPI and SOPI.

The Third Position in Focus. For a last case, consider a study by Henk, Helfeldt, and Rinehart (1985) of the third position. The aim of the study was to determine the relative sensitivity of cloze items to information ranging across sentence boundaries. Only 25 subjects were employed (second position) and two cloze passages (conflating variables of first and third position). The two passages (third position) were presented in a normal order and in a scrambled version (along the lines of Chihara, et al., 1977, and Chavez-Oller, et al., 1985). The relevant contrast would be between item scores in the sequential versus scrambled conditions. Provided the items are really the same and the texts are not different in other respects (i.e., in terms of extraneous variability stemming from first and/or second positions, or unintentional and extraneous adjustments between the scrambled and sequential conditions in the third position).

That is, the tests must not be too easy or too difficult (first position) for the subject sample tested (second position), or, alternatively, that the subject sample does not have too little or too much knowledge (second position) concerning the content (supplied by the first position) of one or both texts, the design at least has the potential of uncovering some items (third position) that are sensitive to constraints ranging beyond sentence boundaries. But does it have the potential for turning up all possible constraints of the type? Or even a representative sampling? Hardly, and there are many uncontrolled variables that fall to the first and second positions that may contaminate the outcome or prevent legitimate contrasts between the sequential and scrambled conditions from showing up

even if they are really there.

In spite of this, the researchers conclude that cloze items don't do much in the way of measuring intersentential constraints. It does not seem to trouble them that this amounts to implying that they have proved that such items are either extremely rare or do not exist at all anywhere in the infinitude of possible texts. This comes near to claiming a proof of the theoretically completely general null hypothesis--that no contrast exists anywhere because none was observed here. This is never a legitimate research conclusion. Anyone can see the difficulty of the line of reasoning if we transform it into an analogous syllogism presented in an inductive order:

> Specific case, first minor premise: I found no gold in California.
> Specific case, second minor premise: I searched in two (or n) places (in California).
> General rule, or conclusion: There is no gold in California.

Anyone can see that any specific case of a similar form will be insufficient to prove any general rule of a similar form. This is not a mere question of statistics, it is a question of a much deeper and more basic form of logic.

CONCLUSION

Therefore, for reasons made clear with each of the several examples with respect to each of the three perspectives discussed, for language testing research and development to be optimally interpretable, care must be taken by researchers to control the variables of whichever of the two positions are not in focus in a particular application of any given test. In the end, in response to Jagjeet Singh (of the International Islamic University in Selangor, Malaysia) who commented that she'd have liked to get more from the lecture version of this paper than she felt she received, I have two things to say. First, that I am glad she said she wanted to receive more and flattered that "the time", as she said, "seemed to fly by" during the oral presentation (I had fun too!), and second, I hope that in years to come as she and other participants reflect on the presentation and the written version they will agree that there was even more to be enjoyed, reflected upon, understood, applied, and grateful for than they were able to understand on first pass. As Alderson correctly insists in his abstract, the study of language tests and their validity "cannot proceed in isolation from developments in language education more generally" (apropos of which, also see

Oller and Perkins, 1978, and Oller, in press). In fact, in order to proceed at all, I am confident that we will have to consider a broader range of both theory and research than has been common up till now.

REFERENCES

Bachman, Lyle. 1988. Problems in examining the validity of the ACTFL oral proficiency interview. Studies in Second Language Acquisition 10:2. 149-164.

Brown, James Dean. 1988. Understanding Research in Second Language Learning: A Teacher's guide to statistics and research design. New York: Cambridge University.

Burks, Arthur W., 1958. Collected Writings of Charles S. Peirce, Volumes VII and VIII. Cambridge, Massachusetts: Harvard University.

Chavez-Oller, Mary Anne, Tetsuro Chihara, Kelley A. Weaver, and John W. Oller, Jr. 1985. When are cloze items sensitive to constraints across sentences? Language Learning 35:2. 181-206.

Chihara, Tetsuro, John W. Oller, Jr., Kelley A. Weaver, and Mary Anne Chavez-Oller. 1977. Are cloze items sensitive to constraints across sentence boundaries? Language Learning 27. 63-73.

Chomsky, Noam A. 1979. Language and Responsibility: Based on Conversations with Mitsou Ronat. New York: Pantheon.

Eisele, Carolyn, ed. 1979. Studies in the Scientific and Mathematical Philosophy of Charles S. Peirce. The Hague: Mouton.

Fisch, Max, et al., eds., 1982. Writings of Charles S. Peirce: A Chronological Edition. Volume 1. Indianapolis: Indiana University.

Hartshorne, Charles and Paul Weiss. 1931-1935. Collected Papers of Charles Sanders Peirce. Volumes 1-6. Cambridge, Massachusetts: Belknapp of Harvard University.

Henk, William A. John P,. Helfeldt, and Steven D. Rinehart. 1985. A metacognitive approach to estimating intersentential integration in cloze tests. In National Reading Conference Yearbook 34. 213-218.

Kamil, Michael L., Margaret Smith-Burke, and Flora Rodriguez-Brown. 1986. The sensitivity of cloze to intersentential integration of information in Spanish bilingual populations. In National Reading Conference Yearbook 35. 334-338.

Maher, Brendan A., Theo C. Manschreck, Cecily C. Weinstein, Margaret L. Schneyer, and Rhoda Okunieff. 1988. Cloze analysis in schizophrenic speech: scoring method and rater's education. Perceptual and Motor Skills 67:3. 911-918.

Moore, Edward C., et al., eds., 1984. Writings of Charles S. Peirce: A Chronological Edition. Volume 2. Indianapolis: Indiana University.

Nagel, Ernest. 1959. Charles Sanders Peirce: a prodigious but little known American philosopher. Scientific American 200. 185-192.

Oller, John W., Jr. 1970. Transformational theory and pragmatics. Modern Language Journal 54. 504-507.

Oller, John W., Jr. 1979. Language Tests at School: A Pragmatic Approach. London: Longman.

Oller, John W., Jr. 1989. Language and Experience: Classic Pragmatism. Lanham, Maryland: University Press of America.

Oller, John W., Jr. 1990. Semiotic theory and language acquisition Invited paper presented at the Forty-first Annual Georgetown Round Table on Languages and Linguistics, Washington, D.C.

Oller, John W., Jr. and Kyle Perkins. 1978. Language in Education: Testing the Tests. London: Longman.

Oltman, Philip K., Lawrence J. Stricker, and Thomas S. Barrows. 1990. Analyzing test structure by multidimensional scaling. Journal of Applied Psychology 75:1. 21-27.

Read, John A. S. 1982. Review of Language Tests at School.

John W. Oller, Jr. London: Longman, 1979. RELC Journal 13:1. 100-107.

Schumann, John. 1983. Art and science in second language acquisition research. In Language Learning, Special Issue: An Epistemology for the Language Sciences, ed. by Alexander Guiora, volume 33.5. 49-76.

Skehan, Peter. 1989. Individual Differences in Second Language Learning. New York: Edward Arnold.

Standing, Lionel G. and Donald Gorassini. 1986. An evaluation of the cloze procedure as a test for plagiarism. Teaching of Psychology 13:3. 130-132.

Zinkhan, George M, William B. Locander, and James H. Leigh. 1986. Dimensional relationships of aided recall and recognition. Journal of Advertising 15:1. 38-46.